DOCUMENT RESUME

ED 157 921 TM 007 268

AUTHOR Meguro, Barbara T.
TITLE Evaluating an Instructional Program's Potential for Producing Student Proficiencies and Identifying Implementation Practices Affecting Program Delivery.
PUB DATE May 78
NOTE 24p.; Paper presented at the Annual Meeting of the California Society of Educational Program Auditors and Evaluators (San Diego, California, May 11-12, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; Course Objectives; *Curriculum Evaluation; *Evaluation Methods; *Instructional Programs; *Program Development; Program Effectiveness; *Success Factors; Teaching Techniques
IDENTIFIERS *Domain Referenced Tests

ABSTRACT

The curriculum evaluation method presented in this paper is focused on determining how an instructional program works in an attempt to explain its instructional accomplishments. The design of the evaluation separates the determination of the instructional program's potential for producing student proficiencies from the evaluation of implementation practices that affect program delivery. The method has been operationalized in conjuction with the continued, naturalistic classroom use of several instructional programs. Program-referenced tests are the key component by providing a sensitive dependent measure to isolate proficiencies gained and to examine the effects of implementation variables. Of particular concern are implementation variables that are under the control of school personnel since identification of these variables has the potential of improving instruction. Since the method is aligned with current inquiries into schooling effects, the results not only provide a specific evaluation but also contribute to the knowledge about the factors that significantly influence school learning. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EVALUATING AN INSTRUCTIONAL PROGRAM'S POTENTIAL FOR PRODUCING STUDENT PROFICIENCIES AND IDENTIFYING IMPLEMENTATION PRACTICES AFFECTING PROGRAM DELIVERY*

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Barbara T. Meguro

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Barbara T. Meguro
SWRL Educational Research and Development

ED157921

Abstract

The curriculum evaluation method presented in this paper is focused on determining how an instructional program works in an attempt to explain its instructional accomplishments. The design of the evaluation separates the determination of the instructional program's potential for producing student proficiencies from the evaluation of implementation practices that affect program delivery. The method has been operationalized in conjunction with the continued, naturalistic classroom use of several instructional programs. Program-referenced tests are the key component by providing a sensitive dependent measure to isolate proficiencies gained and to examine the effects of implementation variables. Of particular concern are implementation variables that are under the control of school personnel since identification of these variables has the potential of improving instruction. Since the method is aligned with current inquiries into schooling effects, the results not only provide a specific evaluation but also contribute to the knowledge about the factors that significantly influence school learning.

---

* Paper presented at the annual meeting of the California Society of Educational Program Auditors and Evaluators, San Diego, May 11-12, 1978.

TM 007 268

EVALUATING AN INSTRUCTIONAL PROGRAM'S POTENTIAL FOR PRODUCING STUDENT PROFICIENCIES AND IDENTIFYING IMPLEMENTATION PRACTICES AFFECTING PROGRAM DELIVERY*

Barbara T. Meguro
SWRL Educational Research and Development

Curriculum evaluations of specific instructional programs are frequently conducted in hopes of finding that acceptable levels of student achievement have been gained and that the effects can be expected to be replicated in other years or other educational settings. Too often, however, the reasons for either the success or failure of the instructional program in a particular situation cannot be explained because the evaluator's focus has been solely on measures of the final effects. According to Raths (1978), however, in his editorial, "Encouraging Trends in Curriculum Evaluation," "evaluators are increasingly interested in learning *how* a particular curriculum program works rather than solely 'proving' various claims to be true (p. 245)." In trying to identify the variables that explain program effects, as Raths states later, the focus is changing from variables such as sex, race, and SES levels, important as they are, to instructional variables "that can be manipulated by teachers and administrators (p. 245)."

In this paper, a process for systematic evaluation of an instructional program over a period of several years is presented. The evaluation process has evolved in conjunction with SWRL's concern for successful

---

*Paper presented at the annual meeting of the California Society of Educational Program Auditors and Evaluators, San Diego, May 11-12, 1978.

classroom use of the outcome-referenced instructional programs developed by SWRL.

SWRL's programs are developed and revised through classroom tryouts, so that, by the time the program is published, the program has already been examined under tryout conditions (Scott, et al., undated) in a "formative" evaluation and revision cycle. The evaluation process described in this paper, therefore, has the purpose of evaluating the potential of the final version of the program under naturalistic classroom use conditions. In this process, the goal is to explain, not only present, the results. The evaluation seeks to identify and determine the significance of the factors that affect the results, with particular interest in those variables that are under the control of school personnel. Since this process is aligned with some of the current research into schooling effects and evaluation research, the results not only provide a specific evaluation but also contribute to the ongoing research into the factors that significantly and consistently influence school learning. To the extent that the findings are explained with variables that occur in other instructional programs, the recommendation of Cronbach (1963) will be fulfilled:

> Insofar as possible, evaluation should be used to understand how the course produces its effects and what parameters influence its effectiveness... Hopefully, evaluation studies will go beyond reporting on this or that course, and help us to understand educational learning.[1]

The instructional programs that have been evaluated by the method described in this paper have certain characteristics. The programs have cognitive learning in various subject areas as their outcomes, and these programs are structured such that all students use identical

or similar materials. That is, the specific student materials are always provided, not just the lesson plans. The scope of each program is fairly extensive, such as a multiyear program in a basic skills area for all primary or elementary grades. The method has been and is being applied while these programs are being used on a large scale in naturalistic school settings, so no experimenter control is being applied during the use of the programs. Further, the evaluation data are being collected throughout the year, not *ex post facto*.

Overview

The process described in this paper evaluates the potential of an instructional program separately from the implementation of that program. The "potential" of the program refers to the learner proficiencies that are attained by expending certain levels of resources with various types of student populations under naturalistic classroom settings. The description of the potential of the program is used by decisionmakers to place a value on the program itself and determine its future. Once the potential of the program has been accepted, the focus shifts to implementation factors that affect the delivery of the program. Data collected previously (while establishing the program's potential) are analyzed to determine the importance of various implementation factors. These analyses seek to provide an explanation of how the program works and thereby identify the implementation variables to be monitored and evaluated in successive years of program use.

Phase I: Determining the Potential of the Instructional Program

The credibility of an instructional program lies in its ability to produce direct learning that is of value to the program adopter. One of the primary concerns in the evaluation of a program, then, is to establish the direct effects of the program. Therefore, the initial phase of the evaluation centers on the potential of the instructional program under various instructional settings. The information necessary to determine the potential of a program include pretest and posttest data on tests that are directly referenced to that instructional program, practices of school personnel responsible for implementing the program, characteristics of the student population, and, if desired, scores on more generalized measures such as standardized achievement tests.

The key to the measurement aspect of the process is the development and use of program-referenced test items in order to detect the proficiencies learned directly from the program. The first and obvious reason to develop program-referenced tests is to demonstrate the proficiencies for which the program can take credit. The personnel responsible for decisions about the program should be provided with accurate information about the program's delivery in order to make value judgments about it. Program-referenced test data are needed to verify whether or not the stated "objectives" of the program are being achieved by most students. A second reason for developing program-referenced tests is to provide a sensitive dependent measure to use later in Phase II when investigating the effects of implementation factors in order to discover how the curriculum program works.

The development of program-referenced test items is described elsewhere in detail (Hanson, Behr, & Meguro, 1977), but can be described briefly as follows. First, the instructional materials are analyzed in depth to identify the outcomes expected to be mastered by all students by the end of each instructional segment. The analysis yields precise domain descriptions of the outcomes taught in each instructional segment. Note that this method of defining domains (outcomes) does *not* utilize either the developer's intended objectives nor judgments about the content that "should be covered" by a program in a particular subject area. Second, item pools are then generated for each outcome of each program segment. Third, the test items are then arranged so that they will be of practical utility to the participants as well as providing data for program evaluation. The items for each program segment are grouped together into a subtest and the subtests are assembled into one test booklet so that the pretest can be used as a placement aid (Hanson, Meguro, & Bailey, 1978). The same test can be administered as a posttest to provide information for planning the next school year. Additionally, individual, longer tests are usually provided for each program segment to be used when a student completes that segment. The scores from these attainment tests, as they are called, feed into an information/reporting system to credit accomplishments and to keep administrators informed about progress during the school year.

The program-referenced pretests and posttests are administered so that each student responds to the items from the instructional segment being used by that student as well as some items from the previous segment and some items from the successive segment. Preferably the

items are also administered to some students at various points of the instruction to demonstrate that each item is learned after specific instruction on that task or, if the skill is cumulative, that proficiency increases with additional instruction.

From the test data, the outcomes that the program can take credit for teaching can be identified. Outcomes that were not mastered by most students after instruction are eliminated. (Usually these outcomes are moved to the next program segment where the cumulative practice is sufficient for mastery.) Then, because the test items and the program are congruent, the pretest-posttest gains can probably be attributed to the program. To check this claim, comparisons are made between the posttest scores of students who receive each program segment, students who received a portion of it, and students who did not receive it. Decreasing scores for the three groups, respectively, provide very convincing evidence of the effects of the program (Hanson, Schutz, and Bailey, 1977). When data are used from many classes that had varying concurrent instruction, the evidence is overwhelming in favor of the program being able to take credit for the outcomes, even though a true experimental design was not employed. Considering the great difficulties of conducting controlled experiments in schools, this method of establishing program effects is appealing above and beyond the appeal of its straightforward, analytic approach to verifying expected results.

Standardized test data may also be presented in the evaluation in an effort to determine the generalizability of results, but the data must be accompanied by an analysis of the correspondence between the

outcomes of the instructional program and the content of the standardized test. The reader is referred to examples of work being done on the correspondence between reading texts and content of standardized tests by Jenkins and Pany (1976) and Armbruster, Stevens, and Rosenshine (1977). Without a content analysis, the simple acceptance of the standardized test data is a *substitute of comparative data* that is usually explained mostly by home background factors (Coleman, et al., 1966; Jencks, et al., 1972) *for a value decision* about the content of an instructional program.

The other aspects of determining the potential of a program are identified by relating the proficiencies attained to the resources used and the characteristics of the student population. The items designed to determine the resources spent include items about the program materials used and the time spent by teachers, aides, tutors, and/or administrators for initial training, teaching, preparation, inservice, and monitoring. The items about the student population request data about demographic and biosocial charactieristics. The items are usually put into a teacher questionnaire and a school questionnaire.[2]

From all the data collected, a descriptive estimate of the potential or "cost-effectiveness" of this program for various types of student populations can be made, although the "cost" may be expressed in terms of resources spent, not dollars. In the interest of equality of educational opportunity, it can be determined from this data whether or not the program produced the same pupil attainment by using the same resources regardless of the characteristics of the student populations.

The information about the potential of an instructional program, when generated by the process described, can be presented during the first two or three years of program use to the personnel responsible for decisionmaking about the program. For examples of such reports, the reader is referred to Meguro (1977) and Anastasiow, Hanson, & Behr (1978). The decisionmakers will attach values to the information, which completes the evaluation of the program's potential for producing student proficiencies.

Phase II: Evaluating the Implementation

Once the potential of an instructional program has been established and a decision has been made to continue its use, the attention of the evaluator turns to the implementation of that program. It is during the evaluation of the implementation that goals can be set with a high degree of confidence and the evaluation can be made in terms of these long range objectives.

The data collected while determining the potential of the instructional program are analyzed to determine estimates of the importance of various teacher and administrator practices on the amount of the instructional program completed. The investigation structures the findings into variables that are controlled by staff. These variables are used as the basis of training, planning, and evaluation in successive years. That is, once it is established that use of the program does produce learning, the focus can shift to implementation variables that affect how much the program is used and, by direct inference, how much is learned. This shift relieves the teacher of the frightening situation of being directly responsible for student scores, and instead makes the teacher responsible for practices such as placement, grouping, use of test results, and most importantly, delivery of the instructional program.

Some preliminary data gathered about SWRL's programs indicate that practices of teachers vary widely, and that these practices affect the amount of the instructional program that is delivered. Delayed start-up at the beginning of the year is a practice that,

a priori, results in a loss of time available for program use. Start-up patterns have been shown to vary widely among program users (Behr, 1978; Behr & Hanson, 1977; Anastasiow et al., 1978). The underplacement of pupils into program sections already mastered is another practice that detrimentally affects the amount of appropriate program delivery (Hanson et al., 1978). Further, the amount of time per day and per week devoted to the program is yet another practice that affects program delivery (Hanson & Schutz, 1975; Hanson & Schutz, 1976). It has also been noted that pacing may be slower than necessary at the beginning of the year, as indicated by the greater amount of time devoted to the initial program segment than for successive segments (Anastasiow et al., 1978). As schooling effects research continues, additional variables to be considered by the evaluator of an instructional program will undoubtedly emerge.

Operationally, the evaluation of the implementation can be carried out through the use of an information/reporting system. The system may include goals set during planning phases, allocation of staff, number of students using the program, placement test results, placement decisions, start-up patterns, periodic reporting of amount of instruction completed, amount of time spent on the program, and scores on the attainment tests taken by students at the completion of each program segment. This information allows for trouble-shooting during the year, comparisons longitudinally from year to year, and, when demanded, a systematic and reasonable way to evaluate the staff who are implementing the program.

The dissemination of the results of the analyses to the educational research community can be a valuable by-product of the evaluation process. The findings, because they are explanatory in nature, have the potential of increasing our understanding of learning, not just evaluating a single program.

Discussion

In response to the theme of the 1978 CSEPAE meeting, *What's happening to the value of evaluation?*, this discussion will focus on the value decisions being made in the evaluation process that has been proposed in this paper.[3] The process itself is summarized in Figure 1, and the steps are discussed below.

Step 1. The first step, adoption of a program, is assumed, and the evaluation process begins after the particular program has been selected. There are any number of reasons for which the program could have been adopted. Context and input evaluations (Stufflebeam, 1971) have been recommended as a careful procedure for selecting a program, but in this paper, no assumptions are made about the method used for program adoption.

Step 2. The evaluation process begins with the development of instruments to determine the potential of the instructional program, namely the program-referenced tests and the teacher and school questionnaires. In the development of the tests, no value decisions are made by the test developer about the content. Instead, the test developer analyzes the program materials and designs the test based on the practice provided in the materials themselves. In the development of the questionnaires about program use and the student population, the developer makes some value judgments when deciding which factors will be investigated, but the decisions are all in answer to the following questions: "What resources were expended for this program?" "What factors characterize the student population?"

The real value decisions being made at this stage are the decisions *not* to collect various other data. The "other" data not being collected fall into three categories: other potential primary effects, comparative effects, and secondary/tertiary effects. The three categories are discussed below.

The reason for focusing on program-referenced tests, not other measures, as evidence of primary effects is the growing evidence in the literature that students learn what they have an opportunity to learn, not something else (Behr, 1978; Wilson, Cohen & Begle, 1968-72; Walker & Schaffarzick, 1974; Buchanan & Milazzo, 1978). Program-referenced tests will accurately reflect the primary effects of the program. Note that there is no reason to exclude items that require a generalization of a skill taught in the program so long as a high level of proficiency is attained by nearly all students after instruction. However, a specific instructional program should *not* be expected to dramatically affect persistent traits as measured on norm-referenced tests, at least not until instructional technology has developed beyond the current stage which, at best, consistently produces student proficiencies on specific content areas. Any standardized test used in the evaluation should be analyzed to determine whether there is any congruence to the instructional program and therefore whether the test is appropriate for the evaluation.

The second type of data not necessarily collected in this method are comparative data on other competing programs. Comparative data may still be the "ideal" research vehicle, but the practical constraints of carrying out controlled studies, or even having competing programs operating simultaneously, are substantial. One possibility, program-fair

testing, provides the "potential" of each competing program by testing all students on the outcomes of all competing programs and providing information about the resources spent and types of student populations involved (Hanson et al., 1978). Program-fair testing, however, is expensive unless program-referenced tests already exist for each program. A second possibility, using norm-referenced, standardized tests to measure growth compared to a national (or perhaps regional or "matched") sample, is methodologically appealing but typically has the drawback of too few items on the relevant content taught by the instructional program to be sensitive to the program's effects. Also, in early grades, the results may be confounded by using different item formats in the test than the students practiced in instruction (Hanson and McMorris, 1977).

Data about secondary and tertiary effects are also not typically collected in the evaluation method presented in this paper. As in the case of comparative data, there is nothing "wrong" about collecting data on the side effects or unintended outcomes, but this data collection could draw the attention of the evaluator away from the primary effect of student learning. Also, the added data collection may be expensive and it may overburden the participants, teachers in particular, to the extent of creating dissatisfaction with the program. Any major side effects will probably become evident to the evaluator, anyway, by unsolicited remarks from teachers or by extremely large or small amounts of program use. Otherwise, just as unintended learning is rare, it seems that major unintended secondary or tertiary effects will be infrequent.

Step 3. Data "collection" and reporting are straightforward steps once the instruments are designed, so no value judgments are made except perhaps such things as methods of aggregating data, explaining data, or organizing the reports. The data collection and reports serve as "process evaluations" (Stufflebeam, 1971) during the first 2-3 years of program use.

Step 4. After a program has been in use 2-3 years, the "potential" of the program should be adequately described as the proficiencies attained by pupils by using certain resources. The school board, funding agency, or group who controls the decision about the program's future then places its values on the "cost effectiveness" information and make a "product evaluation" (Stufflebeam, 1971). It should be noted that at first the results will appear disappointingly small because the excessive claims of many programs have led to unrealistic expectations. Also, it will take some adjustment to accommodate the fact that these results show the minimum levels achieved by virtually all students rather than a composite average that masks the minimum level by including the extra competencies of the above-average students. Nonetheless, the information should be sufficient for evaluating whether or not the program is worth the resources spent, and the information should have been acquired within reasonable cost limitations and with a minimum amount of data collection.

Step 5. Once the decision has been made to continue the program, the continuing evaluation shifts from the program's potential to the implementation factors affecting the program's delivery. The goal here

is to identify practices that are under the control of school personnel that affect the amount of the program completed. No particular value judgments are made in this process except the selection of some practices as the ones to be measured because they are presumed to be the most important ones. The dissemination of the results to the educational research community, Step 5A, does not involve any particular value decisions either.

Step 6. From the analyses of implementation factors that affect program delivery in the previous step, goal-setting can be done along with plans to expend the necessary resources to achieve them. In this planning phase, the values being exercised are the values of the persons setting the goals. The persons involved could be teachers, administrators, evaluators, the school board, the funding agency, or some combination of these personnel.

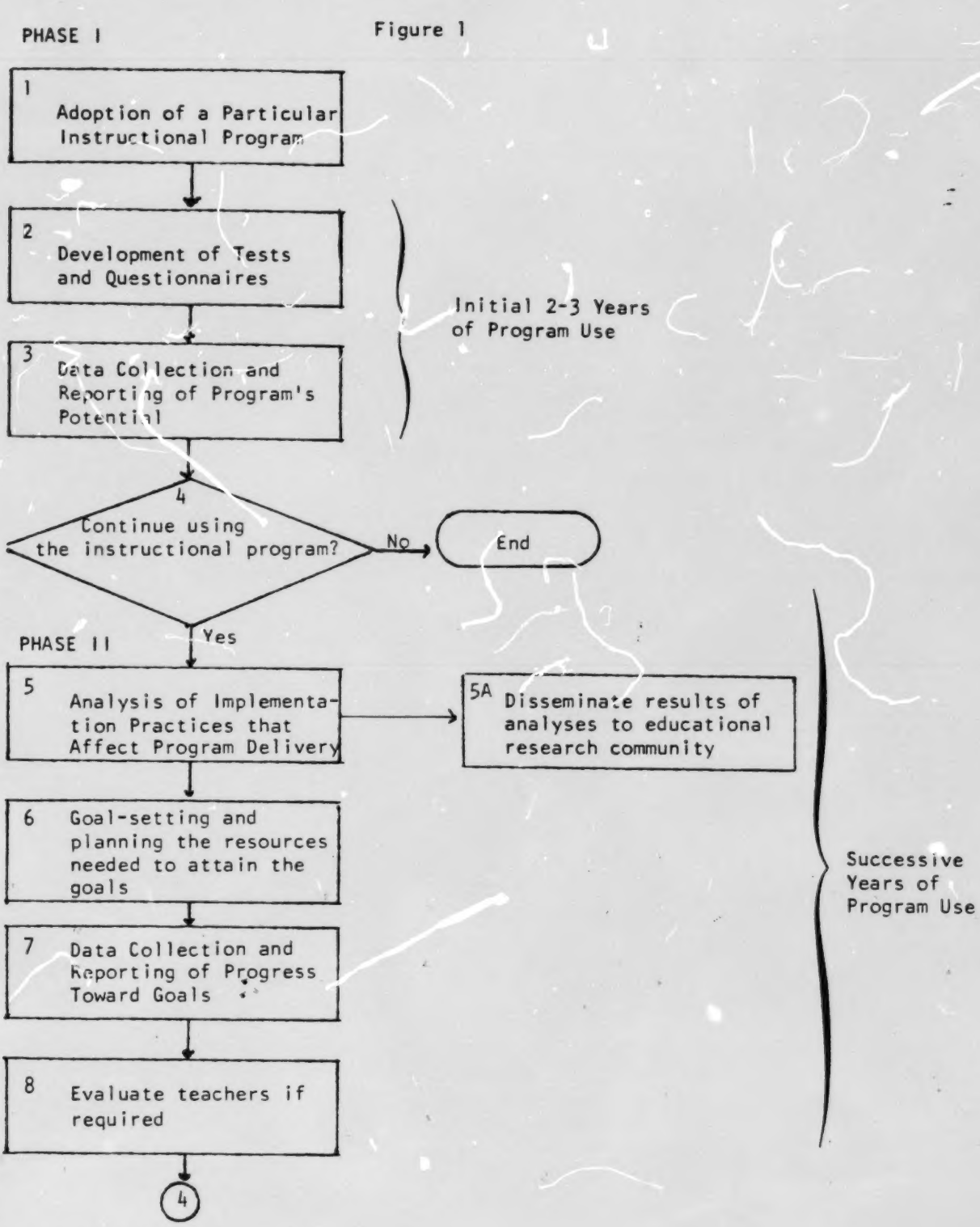
Step 7. Data collection and reporting of progress toward goals, like Step 3, is relatively straightforward with few value judgments involved.

Step 8. If it is required to evaluate teachers, Step 8 provides for evaluating them on the basis of their program delivery relative to the implementation goals set at in Step 6.

Repeat Back to Step 4. The cycle repeats back to Step 4 where a decision is made about the fate of the program. In each successive year, additional information is available on the cost effectiveness of the program and the longitudinal progress made. The value decisions are again made by those person(s) controlling the future of the program.

In summary, the evaluation process presented in this paper primarily serves to provide information to decisionmakers, and it is the decisionmakers who attach value to the information. In the first phase, the information demonstrates the potential of the program to produce student proficiencies using certain resources with certain student populations. If the program is approved for continued use, the focus shifts to the setting of goals and the allocation of the necessary implementation resources to achieve them. An exciting by-product of the second phase is the contribution of the findings to the research literature seeking to explain the relationships between implementation practices and school learning.

Figure 1


PHASE I
1 Adoption of a Particular Instructional Program
2 Development of Tests and Questionnaires
3 Data Collection and Reporting of Program's Potential
Initial 2-3 Years of Program Use
4 Continue using the instructional program?
No
End
Yes
PHASE II
5 Analysis of Implementation Practices that Affect Program Delivery
5A Disseminate results of analyses to educational research community
6 Goal-setting and planning the resources needed to attain the goals
7 Data Collection and Reporting of Progress Toward Goals
8 Evaluate teachers if required
Successive Years of Program Use
4

Notes

Note 1. The quotation is from Page 48 of the reprinted version of "Course Improvement Through Evaluation" contained in Educational Evaluation: Theory and Practice by B.R. Worthen and J.R. Sanders, Worthington, Ohio: Charles A. Jones, 1973.

Note 2. The questionnaires used in Phase I will also include items about implementation practices (grouping, placement, concurrent instruction, etc.) that will be needed in Phase II.

Note 3. The viewpoints presented in this paper, particularly in the discussion, are solely those of the author and not necessarily those of SWRL.

Note 4. Of course, the evaluator must collect other data required by the decisionmaker, particularly any data about goals set at the inception of the project, but in the absence of any such requirements, the method described in this paper does not recommend the use of additional data.

References

Ambruster, B.B., Stevens, R.J., E. Rosenshine, B. Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests. Technical Report No. 26. Center for the Study of Reading, University of Illinois at Urbana-Champaign, March 1977.

Anastasiow, M.A., Hanson, R.A., & Behr, G.E. SWRL/Ginn Communication Skills Program: Descriptive Summary of Quality Assurance Information for 1975-76, Technical Report 61. SWRL Educational Research and Development, Los Alamitos, CA, 1978.

Behr, G.E. Research-Based Considerations for Effective Program Implementation. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Behr, G.E. & Hanson, R.A. Differential Access to Instruction: A Source of Educational Inequality. Paper presented at the annual meeting of the American Educational Research Association, New York, April 1977.

Buchanan, A.D., & Milazzo, P.M. Patterns in K-6 Mathematics Instruction and Achievement. Report 1: Overall Trends. TN 3-78-06. SWRL Educational Research and Development, Los Alamitos, California, 1978.

Coleman, J.S., et al. Equality of Educational Opportunity. Washington, DC: U.S. Government Printing Office, 1966.

Cronbach, L.J. "Course Improvement Through Evaluation." Teacher's College Record, 64, 1963, 672-683.

Hanson, R.A., Behr, G.E., & Meguro, B.T. Domain-Referenced Tests to Support Instructional Programs. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, April 1977.

Hanson, R.A., Meguro, B.T., & Bailey, J.D. Placement Tests to Support Instructional Programs. Paper presented at the annual meeting of the National Council on Measurement in Education, Toronto, March 1978.

Hanson, R.A., & McMorris, R.F. Item Formats and Proficiency Estimation in Domain-Referenced Assessment. TM 5-76-02, SWRL Educational Research and Development, Los Alamitos, California, November 1976.

Hanson, R.A. & Schutz, R.E. The Effects of Programmatic R&D on Schooling and the Effects of Schooling on Students: Lessons from the First-Year Installation of the SWRL/Ginn Kindergarten Program. Technical Report 53. SWRL Educational Research and Development, Los Alamitos, California, June 1975.

Hanson, R.A. & Schutz, R.E. Instructional Product Implementation and Schooling Effects: Lessons from the Second-Year Installation of the SWRL/Ginn Kindergarten Program. Technical Report 56. SWRL Educational Research and Development, Los Alamitos, California, June 1976.

Hanson, R.A., Schutz, R.E., and Bailey, J.D. Program-Fair Evaluation of Instructional Programs: Initial Results of the Kindergarten Reading Readiness Inquiry. Technical Report 57. SWRL Educational Research and Development, Los Alamitos, California, July 1977.

Jencks, E., et al. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972.

Jenkins, J.R., & Pany, D. Curriculum Biases in Reading Achievement Tests. Technical Report No. 16. Center for the Study of Reading, University of Illinois at Urbana-Champaign, November 1976.

Meguro, B.T. Evidence of an Effective Elementary Composition Instruction. Paper presented at the annual meeting of the California Educational Research Association, Los Angeles, November 1977.

Raths, J.D. "Encouraging Trends in Curriculum Evaluation." Educational Leadership, January 1978, 243-6.

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., & Provus, M.M. Educational Evaluation and Decision-Making in Education. Itasca, Illinois: Peacock, 1971.

Scott, R.O., Okada, M., Besei, R., Niedermeyer, F.C., Cronnell, B.A., & Moncrief, M. Development of the SWRL/Ginn Communication Skills Programs: Summary Chronology. SWRL Educational Research and Development, Los Alamitos, CA, undated.

Wilson, J.W., Cohen, L.S., & Begle, E.G. NLSMA Reports. Vols. 1-25 Stanford University: The Board of Trustees of Leland Stanford Junior University, 1968-72.

Worthen, B.R., & Sanders, J.R. Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles A. Jones, 1973.